DOCUMENT RESUME

ED 062 642 CG 007 095

AUTHOR Berlinger, David C.
TITLE The Generalizability of Aptitude-Treatment Interactions Across Subject Matter.
INSTITUTION Far West Lab. for Educational Research and Development, Berkeley, Calif.
PUB DATE Apr 72
NOTE 24p.; Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois, April 3-7, 1972

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Aptitude; *Aptitude Tests; *Instruction; *Learning; Learning Activities; Learning Experience; Learning Processes; *Memory; *Performance; Retention; Teaching

ABSTRACT

This study examined the interactive relations among note-taking (NT), adjunct questions (AQ), and paying attention (PA) by students in lecture situations. The subject pool consisted of 154 students enrolled in basic psychology at the University of Massachusetts. Three aptitude tests were administered to all subjects. These tests measured (1) short term memory, (2) memory for ideas, and (3) memory for sentences. The results were presented in respect to each of the three variables mentioned above. The prevalent pattern was for the memory aptitude to be only slightly positive or negatively correlated with achievement on the criterion test when AQ was the treatment. It was also demonstrated that at the high levels of memory aptitude NT is a very effective instructional strategy for the learner. In addition, this study showed an interaction between the NT and PA treatments. The basic conclusion was that incorporating the aptitudes a subject brings to an instructional situation into a regressive analysis, provides information such that subjects may be assigned to different treatments, each with the greatest potential for individuals with similar aptitude scores. (TL)

FILMED FROM BEST AVAILABLE COPY

N-NPO
CG

ED 062 642

CG 007 095

THE GENERALIZABILITY OF APTITUDE-TREATMENT INTERACTIONS ACROSS SUBJECT MATTER

DAVID C. BERLINER

FAR WEST LABORATORY FOR EDUCATIONAL
RESEARCH AND DEVELOPMENT
1 GARDEN CIRCLE
HOTEL CLAREMONT
BERKELEY, CALIFORNIA
94705

PAPER PRESENTED AT THE MEETING OF
THE AMERICAN EDUCATIONAL RESEARCH ASSOCIATION
CHICAGO, ILLINOIS
APRIL, 1972

CG 007 095

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEH REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

THE GENERALIZATION OF APTITUDE-TREATMENT INTERACTIONS ACROSS SUBJECT MATTER

David C. Berliner*

INTRODUCTION

In previous work (Berliner 1968,1969) claims about the usefulness of note-taking by students in lecture situations were examined and found wanting. Based on that analysis the effects of adjunct questions which could be inserted into lectures appeared to be a technique with considerable promise as an aid to learning from lecture instruction. Moreover, the facilitative effects of such questions has been well established in prose learning studies (Rothkopf, 1965; Rothkopf and Bisbicus, 1967) and fits within a theoretical orientation concerned with mathemagenic behavior (Rothkopf, 1970).

The previous research on learning from lecture examined the effects of note-taking (NT), adjunct questions (AQ), and paying attention (PA), with particular emphasis on aptitude-treatment interactions. The interactions sought were between the three treatments and memory abilities measured in different ways. Berliner (1971) has reported on two of those studies. In the first study it was found that for Ss low in memory ability PA during learning from lecture instruction was as

*Appreciation is acknowledged for the help of many University of Massachusetts staff. Particularly helpful were Jack Lockhead, Angelica Huber Robertson, and Fred J. Dowaliby. Mr. Dowaliby also provided invaluable assistance in the statistical analysis of this study.

effective or more effective than NT. The interaction was disordinal in some analyses. Thus for Ss high in memory ability NT was a superior treatment to PA. The criterion performance was an immediate and delayed short answer criterion test. In some analyses the AQ treatment also interacted with the NT treatment such that for low ability Ss the questions seemed to compensate for deficiencies in memory ability. For high ability Ss NT was a superior treatment to AQ.

In the second study of the series the interaction of NT and AQ was replicated when short term memory was the aptitude measure. The PA-NT interaction was not replicated, but the data did not contradict previous findings. In fact, from examination of the regression lines it still seemed probable that at lower levels of memory ability the differences between NT and PA were negligible. At high levels of memory ability NT appeared to be clearly the superior treatment. Both of the first two studies used Chinese history as lecture material. The third study in the series, described below, examined the interactive relations obtained in the first two studies with a related but different subject matter.

Method

Subjects: The Ss were drawn from an experimental pool consisting of students enrolled in basic psychology at the University of Massachusetts. The Ss signed up for two sessions, on the same afternoon, on successive weeks. Conditions were assigned randomly to time blocks. A total of 154 Ss participated with a mean age of 18.6 years and a range of 17-31 years. All majors from freshman to junior were represented,

though most were at the sophomore level and were liberal arts majors. Forty percent were male and 60% were female. The personal histories of the Ss revealed no differences between treatments for the variables: year in school; major; sex; age; and previous history of the subject matter used in the experimental lectures.

Materials and procedure: Three aptitude tests were administered to all Ss. One of these was an auditory letter-span test, a Short Term Memory Test modified slightly from test MS-3 from the kit of reference tests for cognitive factors (French, 1963). In addition a test of Memory for Ideas, developed by Seibert et al (1967) was used. In this test Ss were asked to listen to a short prose passage taken from a magazine and then record the story. Two stories, one technical and highly factual, the other about the arts, made up the two parts of the test. Central ideas (key words) were scored one point each. It is a test of more than memory, to a small extent the ability to integrate a story would seem to enter into the ability to do well. The third test in this battery was Memory for Sentences. This test required a S to remember complete sentences of various length, and recall them without error immediately after they were spoken. The test sentences were presented to Ss by audio tape along with the other two memory tests. This test, like the Memory for Ideas Test, was developed by the Purdue research group (Seibert, et al, 1967). This test was added to investigate a memory ability calling for more than short term memory storage, related to verbal material, yet less complex than the Memory for Ideas Test. Unfortunately the battery of aptitude tests had to be given after instruction because of time limitations on the use

of the Ss. All tests and accompaning instructions were administered by audio tape.

The lecture material in this study was a 30-minute overview of Indian history, from pre-Alexandrian times to Ghandi, freedom from England and partition with Pakistan. The material, derived from encyclopedias and other standard reference works, was divided into 12 segments of about two and one half minutes duration, each concerned with a specific period of Indian history, or concerned with a specific set of problems or achievements in that history. The materials were designed to be similar to the Chinese history lectures used in the first two studies. Like Chinese history, Indian history is relatively unknown in America and can be described in a manner similar to standard American history courses. The total time of the video taped lecture was 30 minutes. The activities Ss engaged in during the lecture (NT, PA, or AQ) was the independent variable. The dependent variables used were performance on an immediate criterion test (CT-I), and performance on a delayed criterion test, (CT-D), identical to the immediate test but administered one week after instruction. The criterion test consisted of 24 short answer questions culled from every segment of the lecture, and having score values between one and seven points each. A pilot test conducted previous to this study aided in picking items for the criterion test and for use as adjunct questions. Assignment of items for use as adjunct questions during instruction or for testing learning following instruction was made by a random process used on items within each lecture segment. Procedures were followed to insure comparability to the Chiness history lecture and testing.

Results and Discussion

The scoring of the criterion tests and aptitude tests was reliable, with Pearson correlations between scorers of about .90 in separate checks of reliability. Test-retest reliability was also high. For the three treatments the correlations between immediate and delayed tests were .87, .90, and .88. Table 1 presents means, standard deviations and the numbers of Ss per treatment. The results of an F test from an analysis of variance and associated significance levels are also presented for these data. Post hoc analyses using the Newman-Keuls procedure was also completed (column N-K). The results of the Newman-Keuls analysis is presented when a significant between groups difference was present.

Insert Table 1 About Here

A major reversal of previous findings is indicated by the poor test performance of the Ss in the AQ condition in relation to both NT and PA. The NT condition is significantly different from the PA and AQ conditions, which do not significantly differ from each other, as is indicated in the N-K column of Table 1. Previous statements about the overall usefulness of AQ in learning from lecture instruction (e.g., Berliner, 1968) must be scrutinized more carefully in light of the apparent failure of the AQ condition in this study with different subject matter.

An interesting serendipitous finding which occured in all three studies is the fact that the NT condition showed a greater percentage

of a reminiscence effect than the PA or AQ conditions, though again some examples of reminiscence were found in all treatments. The conditions under which all of these studies were performed apparently give rise to a stable reminiscence effect.

Because the aptitude test could potentially have been effected by treatment, since in this study their administration followed by one week the instruction and immediate testing, the aptitude test scores were examined for each treatment. This data is presented in Table 2. No significant differences between treatments were found using simple one way analyses of variance. As noted in Table 2, means and standard deviations for the tests used in each treatment are quite homogenous.

Insert Table 2 About Here

The effects of the treatments used in this study were analyzed by a regression approach. The regression equations used on these analyses are found in Table 3. With regard to the aptitude measured by the Short Term Memory Test no significant departure from paralellism was found for the regression lines used in the analysis of immediate and delayed criterion tests. However, the data suggests that at low levels of this aptitude NT is a more effective treatment than PA. This datum, though not significant, is not in line with the earlier studies and illustrates that the generalization of ATI's to different subject matter areas may be difficult. In fact, this is the only case out of many analyses where the slope of the NT group is negative.

Insert Table 3 About Here

The regression of CT-I on Memory for Ideas, Part 1, did produce a disordinal interaction between the PA and AQ treatments. Paralellism was rejected; F=5.03, df=1/115, p=.03. Homogeneity of variance existed. The application of the modified Johnson-Neyman technique (Dowaliby and Berliner, 1971) described a region of non-significance from 12.91 to 108.91 around a point of non-significance of 22.53 on the X axis. Seventeen Ss, 14.3% of this sample, fell below the region of non-significance. As displayed in Figure 1, for Ss who are below a score of 12.91 points in this memory aptitude, it would be recommended that they pay attention rather than use a treatment incorporating AQ. Confidence in that recommendation is 90%. At the higher levels of this memory aptitude the AQ treatment may be recommended over the PA treatment, though confidence in that recommendation is considerably less than 90%. Though the trends were similar, no significant interactions were noted in the analysis of the regression of CT-D on Memory for Ideas, Part 1.

Insert Figure 1 About Here

Figure 2 displays the regression lines for the analysis of CT-I on Memory for Ideas, Part 2, while Figure 3 displays the regression of CT-D on the same aptitude test. With regard to CT-I the hypothesis of common slope for the PA and AQ regression lines was rejected; F=5.00, df=1/115, p=.03. The assumption of homogeneity of variance was met and the modified Johnson-Neyman technique described a region of non-significance extending from 6.24 to 62.63, around a point of non-

significance at 14.14 on the X axis. Ten Ss, 8.4% of this sample, fell below the region of non-significance. With 90% confidence it may be stated that Ss scoring below 6.24 on this memory aptitude test would benefit more from a treatment in which they pay attention than one which uses AQ. With less confidence one would posit that Ss high in this memory aptitude would benefit more from a treatment incorporating AQ than one which calls for PA.

Insert Figure 2 About Here

In the case of the delayed criterion test, illustrated in Figure 3, an interaction between NT and PA appeared. The hypothesis of common slope for their regression lines was rejected; $F=5.05$, $df=1/93$, $p=.03$. The assumption of homogeneity of variance was not rejected. Use of the modified Johnson-Neyman technique revealed a region of non-significance from -218.38 to 10.77 around a point of non-significance at 5.68 on the X axis. Seventy Ss, constituting 72.2% of this sample, fell above the region of non-significance. Thus for these cases, with 90% confidence, it can be recommended that they would significantly benefit from a treatment requiring NT. The differences between the NT and PA conditions are not significant at the lower aptitude levels.

Insert Figure 3 About Here

While the relationship between the AQ and PA treatments is, in this study, different from previous studies, the relationships between the NT and PA treatments described in Figures 2 and 3 are quite

supportive of previous interpretations. When memory aptitude is high, NT clearly exceeds PA in utility. When memory aptitude is low, those differences are less or even reversed, such that PA is sometimes seen as more effective than NT. Both the immediate and delayed criterion test show almost identical configurations when regressed on Memory for Ideas, Part 2.

The relationships noted with regard to Figures 2 and 3 are quite similar to those obtained when CT-I and CT-D are regressed upon the full scale score for the Memory for Ideas Test. The interaction between AQ and PA was significant ($p < .05$) with regard to immediate and delayed testing, but a region of significance was found only in the case of the delayed criterion test. The interaction between NT and PA failed to reach significance, but the relationship suggested by the regression lines for these two conditions, in this analysis, is supportive of all previous interpretations. Because of the similarity to Figures 2 and 3, no figures have been drawn to illustrate the regression of CT-I and CT-D on the full scale Memory for Ideas Test. However, the regression equations are provided in Table 3.

The use of the aptitude test Memory for Sentences provided no statistically significant support or disconfirmation of other findings. The regression equations for these analyses are provided in Table 3, but no trends were discernable from these data.

SUMMARY

AQ: Previous discussions of the main effects occuring in Study 1, which was designed to investigate the effects of AQ in oral instruction,

must be modified in light of the ATIs which were found in the analysis of three studies and the lack of potency of the AQ treatment noted in this study. Within the subject matter of Chinese history AQ appears to be a useful technique in learning from lecture instruction, provided memory aptitude scores are low. This statement is supported by many analyses and seems to hold particularly for the aptitudes of Short Term Memory and Memory for Ideas. Supporting but non-significant data was also found in the analysis of the Memory for Sentences Aptitude. The prevelant pattern was for the memory aptitude to be only slightly positive or negatively correlated with achievement on the criterion test when AQ was the treatment. It is possible that AQ often function as a memory aid to Ss low in memory aptitude. The typically high correlation between memory aptitude and criterion test score found in the NT treatment (e.g. $r=.72$ between Memory for Ideas, Part 1, and a criterion test in study one) indicates a strong relationship between memory aptitude and criterion test performance after taking notes. But that relationship is vitiated when a S uses AQ (e.g. $r=-.03$ between Memory for Ideas, Part 1, and a criterion test in study one). The lessened dependence on memory for criterion test achievement in the AQ condition was strongly supported in Study 1, partially supported in Study 2, but does <u>not</u> hold in this study of generalizability wherein the AQ treatment was not very powerful at all.

Careful analysis of the Indian history lecture and test items revealed no obvious differences from the Chinese history material. Any attempt at explaining the observed differences between the first two studies and the third study is premature. Should the conditions

be uncovered which in this study differ from the other studies, we may posit that under those conditions AQ are likely to produce higher performance on criterion tests than PA at higher levels of certain memory aptitudes. It should be noted, though, that a significant between groups difference clearly point out the overall efficacy of NT under the conditions of this study, particularly at these higher levels of memory aptitude.

NT: The literature on the behavior of note-taking during lecture has generally pointed out its lack of effectiveness. In the studies reported above there is considerable support for the taking of notes by Ss only if their memory aptitude is high. All three studies, in one analysis or another, demonstrated that at the high levels of memory aptitude NT is a very effective instructional strategy for the learner. In addition, the studies generally describe an interaction between the NT and PA treatments. In most instances the NT treatment has been characterized by a regression line that has a steeply rising slope while the PA treatment has been generally characterized by a relatively flat slope. It can be stated with the confidence inherent in three replications that when memory aptitude is low the advantage of NT over PA is minimum and that sometimes the PA treatment is even superior to the NT treatment at these levels of measured memory aptitude. Figure 4 shows this relationship as reported in study one. Figure 5 shows the relationship uncovered in study 2. The data reported for this study of generalizability over subject matter, as given in Figures 2 and 3 displays similar trends.

Insert Figures 4 & 5 About Here

This interpretation of the data contradicts the folk psychology about one's memory ability and procedures to follow in learning from oral instruction. Nevertheless, a dispassionate psychological analysis of note-taking is quite viable, wherein it is hypothesized that only when memory aptitude is "good" does one possess the ability to accurately store orally transmitted information for the time necessary to transcribe it accurately. Further, only when memory aptitude is "good" can one store and attend to the new information that is being transmitted while transcribing any previously obtained information. When memory aptitude is low these abilities are likely not to be present, and it is as efficient or more efficient for the learner to pay attention to the lecture. Careful analysis of note-taking for Ss of low and high memory aptitude has not been undertaken by this investigator but should be worth examining at some future time.

PA: At low levels of measured memory aptitude, when PA may be as efficient as NT, the AQ treatment has generally shown its greatest potential, though that is only true of studies one and two and is not true in this study of generalizability. At the high levels of memory aptitudes, whether PA is more beneficial than AQ or vice versa, the NT treatment has been shown to be clearly the most beneficial behavior for a learner to engage in during instruction by lecture. Thus, it appears unlikely that PA is ever to be the recommended behavior for a S. However, the design of instruction to incorporate AQ is quite costly in both time and money for most situations. Only when a cost-benefit analysis indicates its utility is it likely to be used in lecture instruction. It is possible that under the typical conditions operating

in college instruction, for example, only the NT and PA treatments are likely to occur and in this case it is the NT treatment that costs the most in energy expended. In this case PA may be recommended for Ss low in measured memory aptitude.

ATIs: Aptitude-trea ment interactions, relying upon the little used statistical technique of Johnson and Neyman (1936) for analysis are not often found in the literature. However, a trend toward analysis of data incorporating the ATI approach and methodology is developing (cf. Berliner and Cahen, in press). This paper points out that there were a number of interesting ATIs in Study 1, and a replication of some of these in Study 2, with confirmation of some (though not all) of these trends in this study of generalizability. These data should go a long way toward dispelling doubts about the robustness of ATIs. In the case of the NT and PA interaction we see support continuing over time (immediate and delayed testing), samples (San Jose State College and University of Massachusetts), and subject matter (Chinese history and Indian history). The AQ and NT interaction was shown to hold over time and samples, but not over subject matter.

While the conditions under which certain ATIs hold or do not hold must be investigated further, and many more aptitudes must be examined for interactive effects, it is noted in conclusion that the ATI approach to instruction has great potential in educational psychology. Clearly, the effects noted in the analyses of variance reported above were inadequate to understand the trends in these data. Further, using a simple effects analysis after finding an interaction with an

analysis of variance would not have produced the precision in locating the points at which confident decisions could be made. Using the aptitudes an S brings to an instructional situation, and incorporating these into a regression approach to analysis, provides information to a decision maker such that Ss may be assigned to <u>different</u> treatments, each with the greatest potential for individuals with similar aptitude scores.

REFERENCES

Berliner, D. C. The effects of test-like events and note-taking on learning from lecture instruction. Unpublished doctoral dissertation, Stanford University, 1968.

Berliner, D. C. Effects of test-like events and note-taking on learning from lecture instruction. Paper presented at the meeting of the American Psychological Association, Washington, D.C. September, 1969. (Issued also as Technical Report Number 2, Center for Educational Research, School of Education, University of Massachusetts, Amherst, Massachusetts. March, 1970.)

Berliner, D. C. Aptitude-treatment interactions on two studies of learning from lecture instruction. Paper presented at the meeting of the American Educational Research Association, New York City, February, 1971.

Berliner, D. C. and Cahen, L. S. Trait-treatment interactions and learning. In Kerlinger, F. (Ed.) The Annual Review of Educational Research. Itaska, Illinois: F. E. Peacock, 1973 (in press).

Dowaliby, F. J., and Berliner, D. C. ANALATI: A computer program for the application of the Johnson-Neyman technique in the case of 1 predictor and 1 criterion variable. Technical Report No. 15, Center for Educational Research, School of Education, University of Massachusetts, Amherst, Massachusetts, November, 1971.

French, J., Ekstrom, Ruth B., and Price, L. A. Manual for kit of reference tests for cognitive factors. Princeton, New Jersey: Educational Testing Service, 1963.

Johnson, P. O. and Neyman, J. Tests of certain linear hypotheses and their application to some educational problems. Statistical Research Memoirs, 1936, 1, 57-93.

Rothkopf, E. Z. Some theoretical and experimental approaches to problems in written instruction. In J. D. Krumboltz (Ed.), Learning and the educational process. Chicago: Rand McNally, 1965. Pp. 193-221.

Rothkopf, E. Z. The concept of mathemagenic activities. Review of Educational Research, 1970, 40, 325-336.

Rothkopf, E. Z., and Bisbicos, E. E. Selective facilitative effects of interspersed questions on learning from written materials, Journal of Educational Psychology, 1967, 58, 56-61.

Seibert, W. F., Reid, J. C., and Snow, R. E. Studies in cine-psychometry II: continued factoring on audio and visual cognition and memory. Lafayette, Indiana: Audio Visual Center, Purdue University, Final Report, December, 1967.

Table 1. Descriptive statistics and summary analysis of the criterion tests in the generalizability study.

		Treatment				
		AQ	NT	PA	F test	N-K
CT-I	M	19.53	25.91	21.07	**	NT > AQ*
	SD	8.98	8.50	8.44		NT > PA*
	N	57	35	62		
CT-D	M	16.86	24.03	18.60	**	NT > AQ*
	SD	8.23	8.59	8.56		NT > PA*
	N	57	35	62		

* = Significant difference between groups at the .05 level of confidence
** = Significant difference between groups at the .01 level of confidence

Table 2. Descriptive statistics for the aptitude tests used in the generalizability study

		Treatments in Generalizability Study		
		AQ	NT	PA
Short Term Memory Test	M	7.30	6.83	7.18
	SD	2.67	1.95	2.34
	Range	1-14	3-11	2-13
Memory for Ideas, Part 1	M	19.54	22.20	20.21
	SD	7.16	5.43	7.19
	Range	0-33	8-34	0-37
Memory for Ideas, Part 2	M	11.93	12.23	13.58
	SD	5.14	5.18	4.39
	Range	2-21	0-23	3-23
Memory for Ideas, Full Scale	M	31.49	34.43	33.79
	SD	10.03	8.38	9.12
	Range	4-52	17-49	10-50
Memory for Sentences	M	2.45	2.23	2.34
	SD	1.90	1.35	1.44
	Range	0-8	0-5	0-8

Table 3. Regression equations and variables used in the generalizability study.

Regressed Variables	Regression Equations		
	NT	AQ	PA
CT-I on Short Term Memory	Y=28.17- .33x	Y=20.73-.16x	Y=14.01+ .98x
CT-D on Short Term Memory	Y=29.00- .73x	Y=18.57- .23x	Y=14.86+ .52x
CT-I on Memory for Ideas, Part 1	Y=13.34+ .57x	Y= 5.56+ .71x	Y=15.87+ .26x
CT-D on Memory for Ideas, Part 1	Y=14.61+ .42x	Y= 4.87+ .61x	Y=13.53+ .25x
CT-I on Memory for Ideas, Part 2	Y=19.34+ .54x	Y=11.40+ .68x	Y=21.82- .06x
CT-D on Memory for Ideas, Part 2	Y=14.35+ .79x	Y=10.77+ .51x	Y=19.03- .03x
CT-I on Memory for Ideas, Full Scale	Y=10.65+ .44x	Y= 2.47+ .54x	Y=16.10+ .15x
CT-D on Memory for Ideas, Full Scale	Y= 7.49+ .48x	Y= 2.85+ .44x	Y=13.58+ .15x
CT-I on Memory for Sentences	Y=24.28+ .73x	Y=15.85+1.50x	Y=16.02+2.16x
CT-D on Memory for Sentences	Y=21.56+1.11x	Y=14.77+ .85x	Y=14.56+1.73x

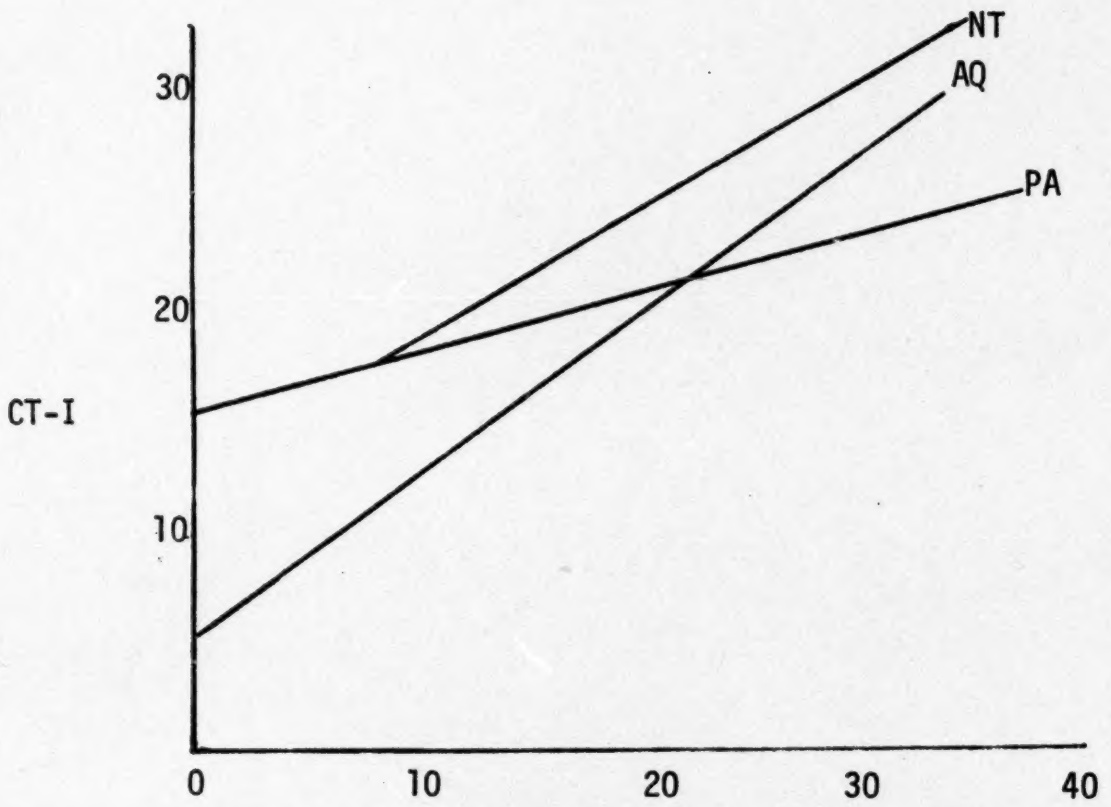

Figure 1. Regression of CT-I on Memory for Ideas, Part 1

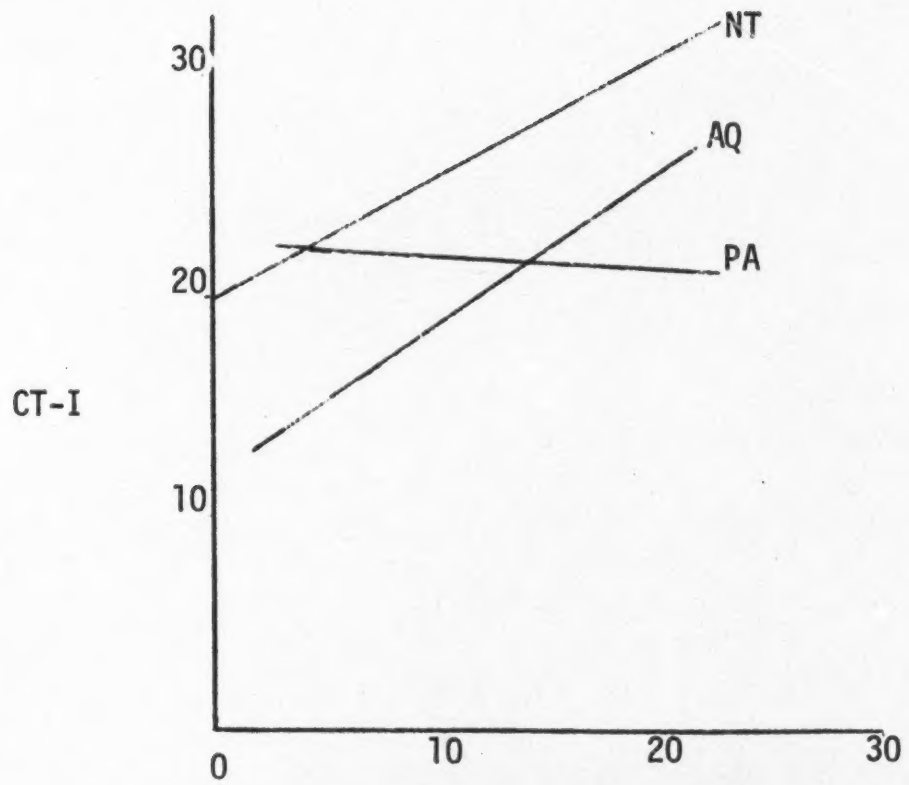

Figure 2. Regression of CT-I on Memory for Ideas, Part 2

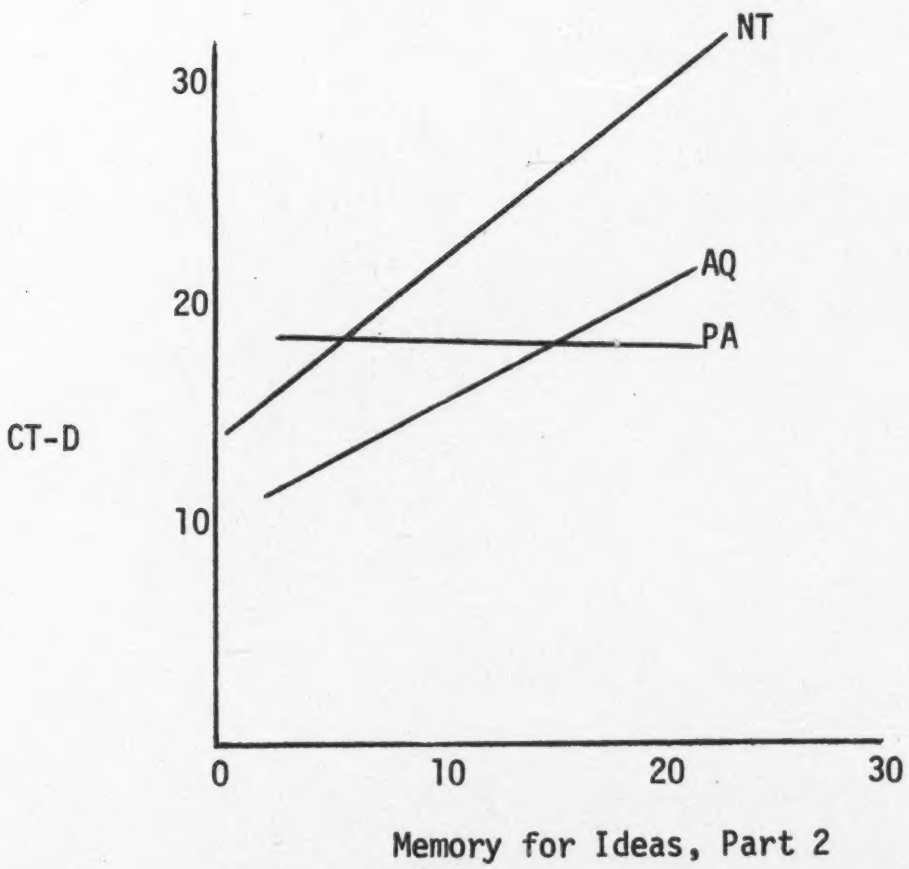

Figure 3. Regression of CT-D on Memory for Ideas, Part 2

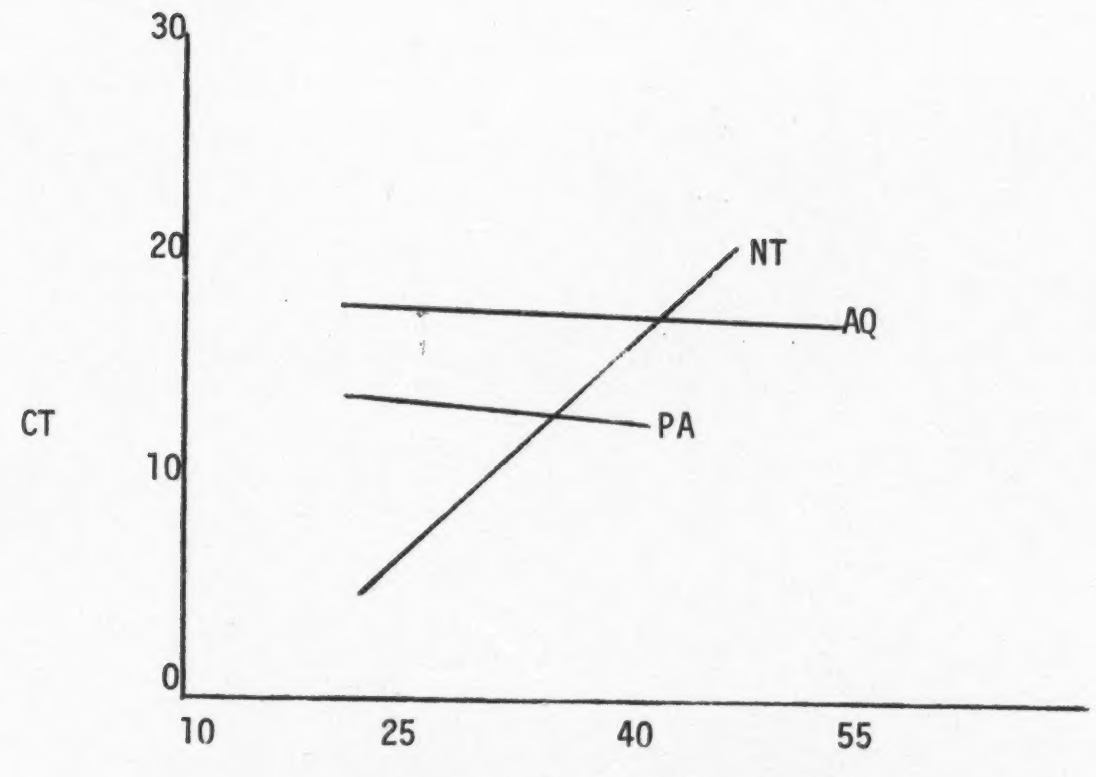

Figure 4. Regression of CT on full scale Memory for Ideas in study one

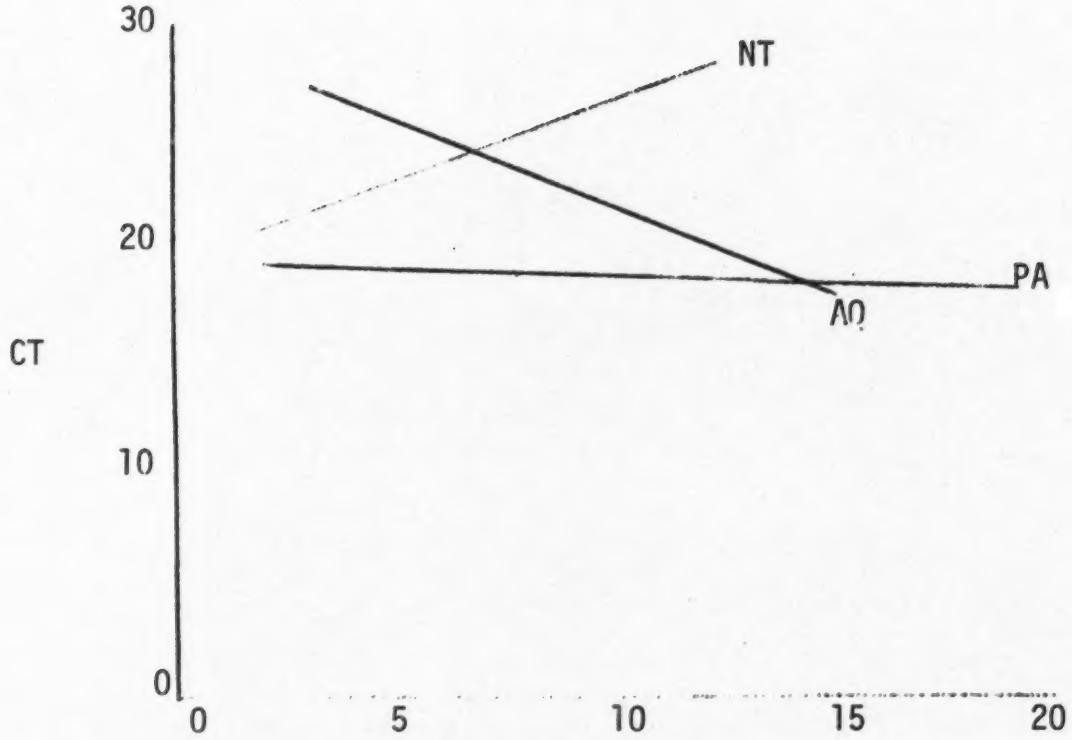

Figure 5. Regression of CT on Short Term Memory Test in Study Two